Competitive Advantage Through Cloud Computing
November 15, 2011

Cloud Computing in the Era of "Big Data"

Business Considerations

James Short
Lead Scientist
Center for Large-scale Data Systems
San Diego Supercomputer Center
UC San Diego
jshort@ucsd.edu

Agenda

- Overview: Will Cloud be as big as the Internet? *(warning label)*
- Drivers *(no one goes there anymore, it's too busy)*
- "Big Data" and Business Analytics *(our previous era of little data)*
  - How much data? How much information?
  - A report on server capacity and performance
  - Workloads
  - M2M
- Business Considerations for Cloud *(time for a reality check)*
  - Who owns cloud in your enterprise? Who is responsible for it?
  - Business value assessment
    - Is cloud strategic (an enterprise investment)? Is cloud "virtualization and data center consolidation" (a cost-focused, business unit IT initiative)?
  - Some Early Lessons: Cloud TOP 10 list (business)

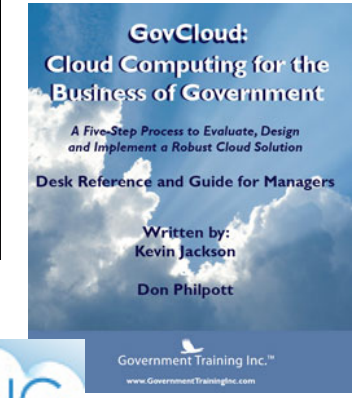
Loud Clouds in the Market

Forbes

Business

9/07/2011 @ 11:54AM |2,999 views

Will Cloud Computing Really Be As Big As The Internet?



"Do Attorneys Know The Difference Between Their SaaS and a Hole In The Ground?"

InsideLegal
www.insidelegal.com

CloudTweaks
PLUGGING INTO THE CLOUD

Top 55 Cloud Computing Enablers Gaining Mind Share

http://www.cloudtweaks.com/2011/03/top-55-cloud-computing-enablers-gaining-mind-share-in-1q-2011/

1. Amazon Web Services (AWS),
   - Elastic Compute Cloud (EC2),
   - Simple Storage Service (S3),
   - Virtual Private Cloud (VPC),
   - Elastic Beanstalk,
   - Elastic Block Store (EBS),
   - CloudFront,
   - Cloud Formation,
   - CloudWatch
2. Salesforce.com
   - Sales Cloud 2 (CRM),
   - Service Cloud 2 (Support),
   - Force.com (Development Platform),
   - Chatter (Collaboration),
   - Heroku,
   - Database.com,
   - Jigsaw,
   - RemedyForce,
   - AppExchange
3. Google Apps (Web based business applications suite)
4. VMWare (Virtualization management)
   - vCloud Director,
   - vCenter,
   - vSphere
5. Citrix – XenServer (Virtualization), VMLogix
6. Rackspace – Cloud Builders / OpenStack, CloudKick
7. HP - 3PAR, Vertica (Data Analytics)
8. Dell – Boomi (AtomSphere – Cloud Integration)
9. Cisco
10. CA Technologies - 3Tera AppLogic (Turnkey Cloud computing platform)

Rapidly evolving business marketplace

…

Vint Cerf, Google:

"No standards"

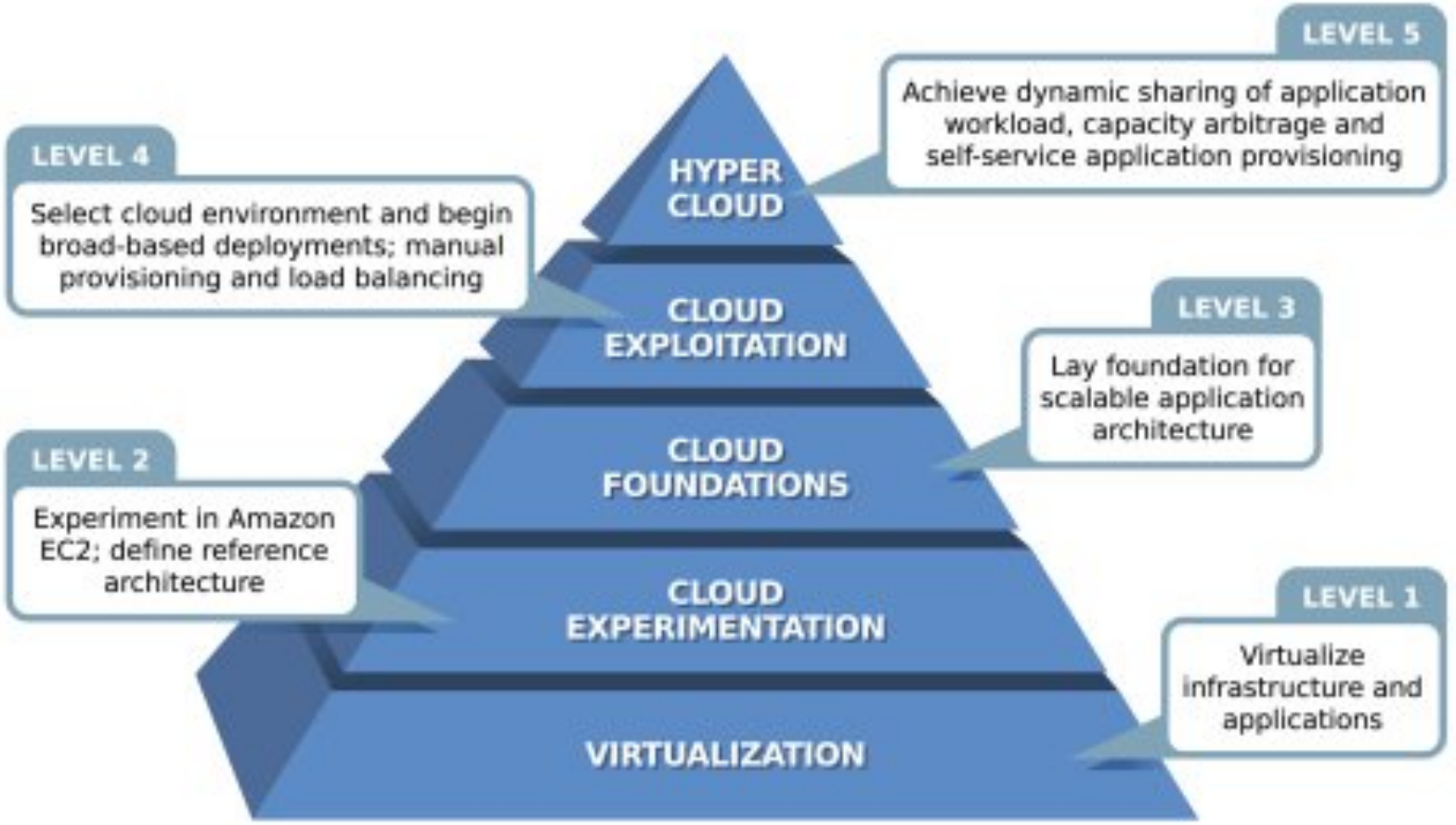
What Is Cloud?

- Infrastructure provided by a service provider to build and operate Internet applications
- New model of consumption, operations and delivery of IT services
- Cloud categories
  - Software as a Service
  - Infrastructure as a Service
  - Platform as a Service
  - Database as a Service

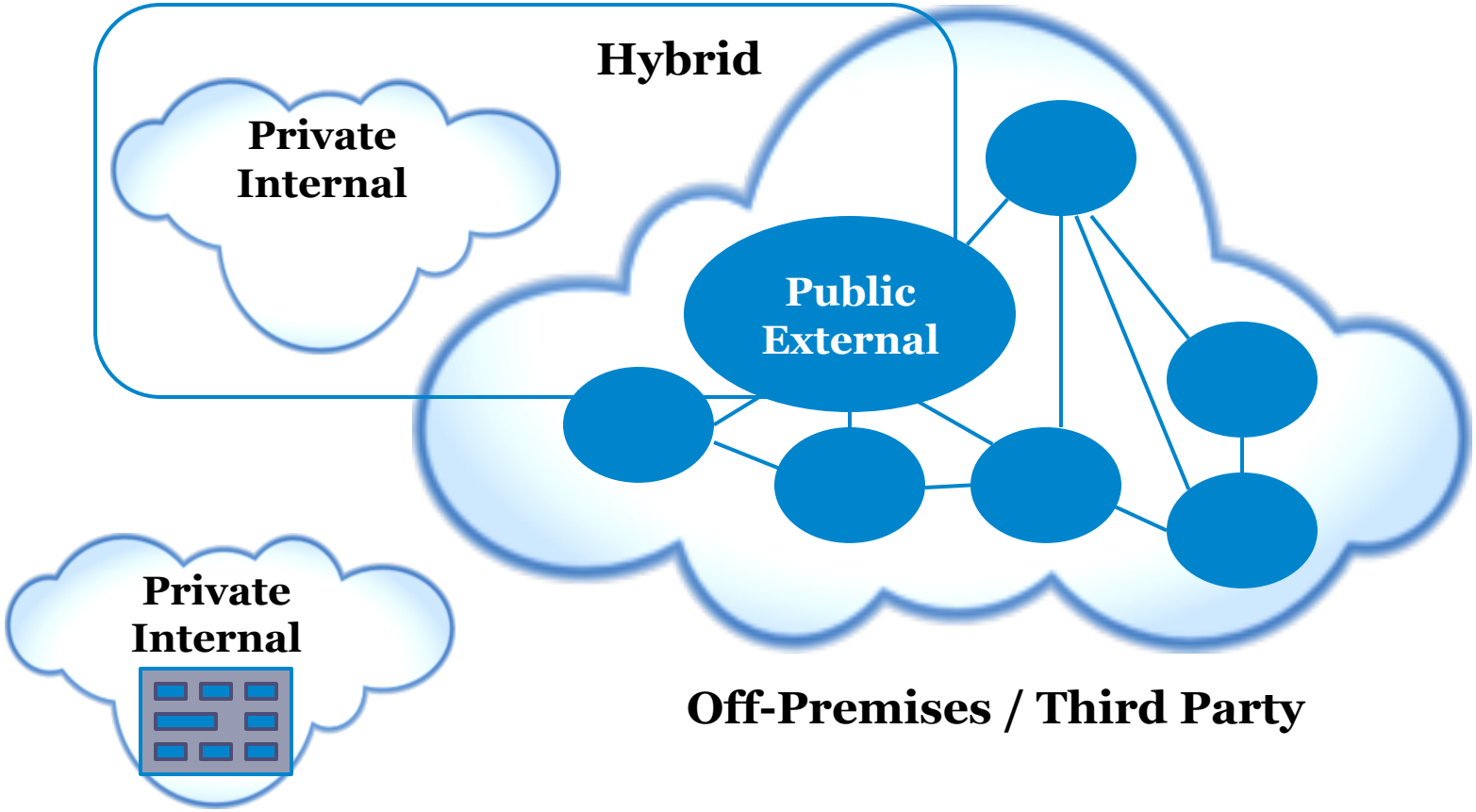
Why Cloud?
Business IT Considerations

- End user focus (ability to directly configure resources)
- Multi-tenancy (shared services)
  - Utilization and efficiency improvements
  - Peak load capacity
  - Centralization of infrastructure components
- Cost
  - Capex converted to operational expense
  - Utility computing pricing (usage based options…)
- Reliability / Scalability / Performance
  - Redundant sites, dynamic provisioning, performance monitoring
- Device and Location independence
  - Users access via web browser anywhere
- Information security / governance
  - Centralization of data, but data is distributed over a wider area and more devices

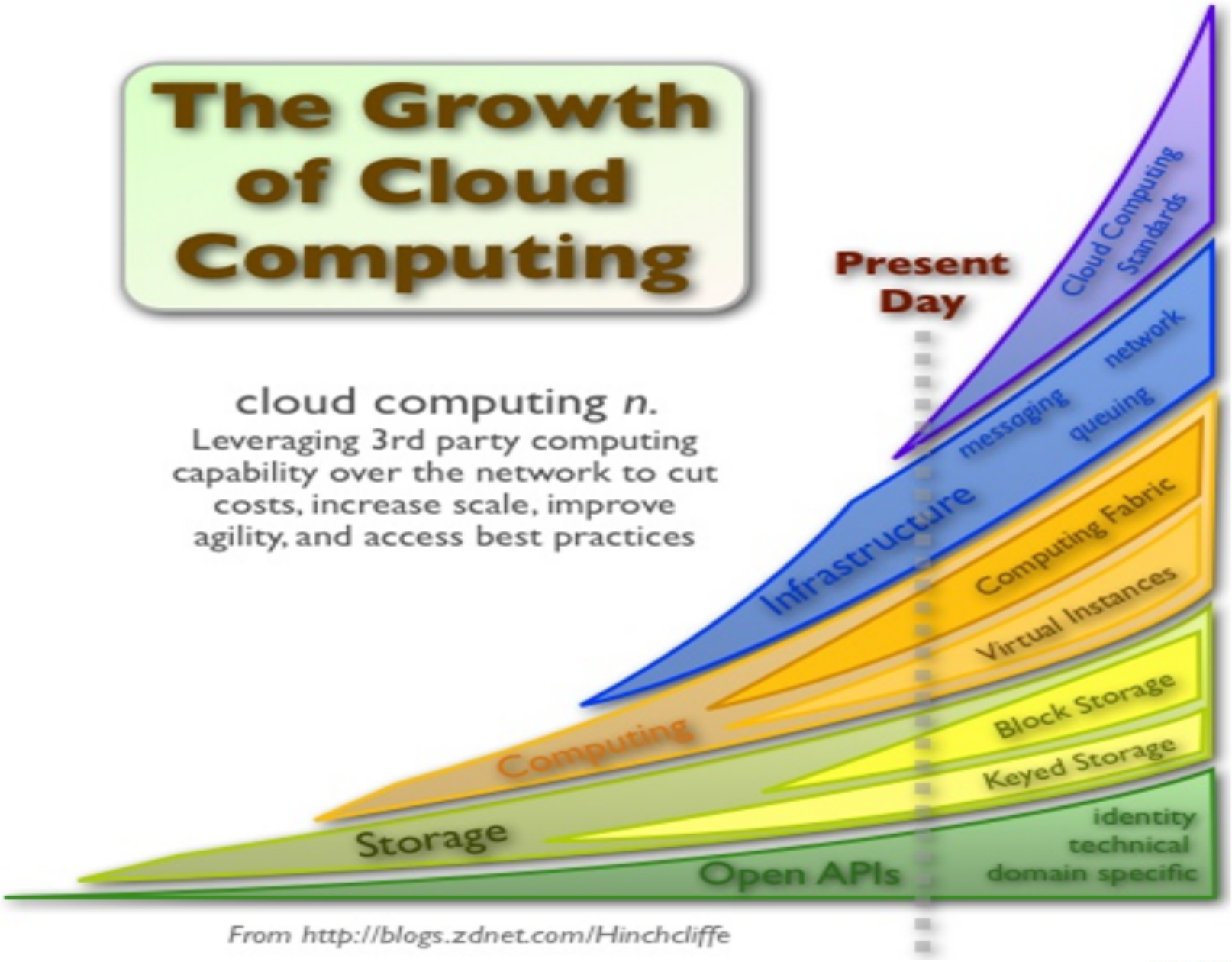
THE CLOUD COMPUTING ADOPTION MODEL

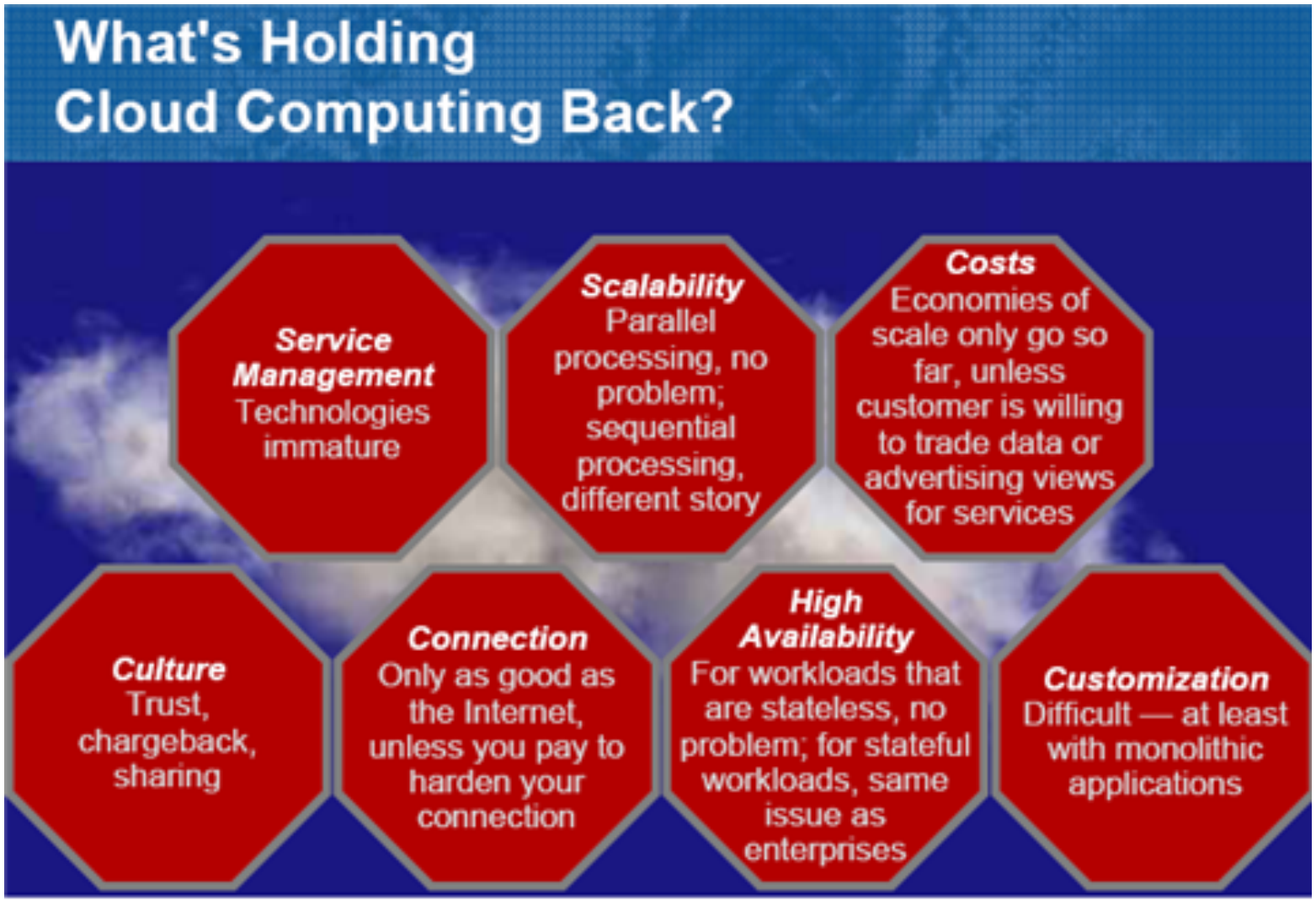


Cloud Deployment



Public Cloud, Community Cloud, Hybrid Cloud, Private Cloud


The Growth of Cloud Computing
cloud computing n.
Leveraging 3rd party computing capability over the network to cut costs, increase scale, improve agility, and access best practices
Present Day
Cloud Computing Standards
Infrastructure
messaging
network
queuing
Computing
Computing Fabric
Virtual Instances
Storage
Block Storage
Keyed Storage
Open APIs
identity
technical
domain specific
From http://blogs.zdnet.com/Hinchcliffe


What's Holding Cloud Computing Back?
Service Management
Technologies immature
Scalability
Parallel processing, no problem; sequential processing, different story
Costs
Economies of scale only go so far, unless customer is willing to trade data or advertising views for services
Culture
Trust, chargeback, sharing
Connection
Only as good as the Internet, unless you pay to harden your connection
High Availability
For workloads that are stateless, no problem; for stateful workloads, same issue as enterprises
Customization
Difficult — at least with monolithic applications
Gartner

"Cloud-ify"

- *"Having been at the heart of a major "private cloud" build out for a large multinational company over the last 3 years, I can offer my opinion that "private clouds" are really nothing more than an analogy for "data center consolidation". The truth of the matter is that large corporations are riddled with legacy technologies which is almost impossible to "cloud-ify". Foundational elements such as virtualization, automated provisioning (which I still think nobody in the vendor space has truly cracked) and streamlined work processes are the enablers of better run, more efficient data centers, but "private clouds" they are not. If an organization must continue to bear the capital costs, fixed not variable, then it is hard to make the statement that this is an "on demand" data center. The only true cloud is the one you can pay for with your credit card. You can't do that internally."*

DRIVERS

Cloud Computing In the Era of Big Data

Big Data – Example Stats

- 30 billion pieces of content were added to Facebook… in the past month
- More than 2 billion videos were watched on YouTube … yesterday
- The average teenager sends 4,762 text messages per month
- 32 billion searches were performed last month … on Twitter
- Netflix accounts for 32.7% of peak US downstream traffic

Sources: Gartner, Netflix, CNN, Sandvine

Satellite Telemetry:
1 PB per day, HD full motion video

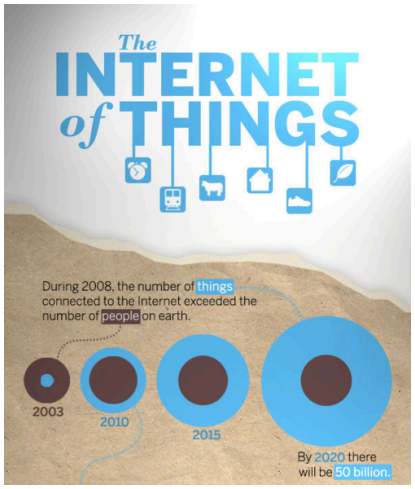
How Much Surveillance Data?
NYT Review of Books says NSA Datacenter will store Yottabytes of data (10^{24} bytes). The Internet is about 1 Yottabyte.

WAL★MART

Real Time Analytics:

- *Everything Wal Mart knows about a customer*
- *Everything our trading partners know about the customer*
- *Everything on the public internet about the customer*
- *Combine it, process it in real time for all customer interactions*
- *And store it for future use*

50 Billion Devices By 2020



Real Time Ad Placement On Social Media Websites

Big Data at Twitter

- Three Challenges
- #1 Collecting Data
  - 7 TB/day, 2+ PB a year
  - 225 GB during this presentation
  - Early DB did not scale, lost data, transitioned to Scribe
- #2 Large-Scale Storage and Analysis
  - How store 7 TB/day (in Hadoop cluster)
  - Analysis at scale (12 billion tweets, 5 mins)
  - Social graph analysis over hybrid DB arrangement

Source: http://www.slideshare.net/kevinweil/big-data-at-twitter-chirp-2010

Big Data at Twitter

- Three Challenges *(continued)*
- #3 Rapid Learning Over Big Data
  - Use PIG
  - Counting Big Data
    - How many requests per day
    - Average latency? 95% latency?
    - Response code distribution per hour
    - Searches per day
    - Unique users searching? Unique queries?
    - Geographic differences for queries?
    - Usage differences for mobile users

Scale

- New concepts and practices that use scale
  - Web 2.0, Collective Intelligence, Crowdsourcing
  - Social media, Internet of Things, Wikinomics
- New technologies, behaviors, cultural forms that are taking advantage of scale
  - Multi-player online games
  - Wikipedia, blogs, community mapping
  - MyLifeBits project at MSR
  - London Travel Planning
    - http://www.tfl.gov.uk/

Scale

- Real Time Business
  - Constant adjustment rather than fixed rules and prices
  - Think of:
    - Numerous networks adjusting weights, prices, offerings, connections
    - New technologies that will automatically summarize information and place it into networks and hierarchies
    - The reach and range of “analytics” expands with data
      - Business, social, cultural

How Much Information Is There Anyway?

Findings from HMI? 2009

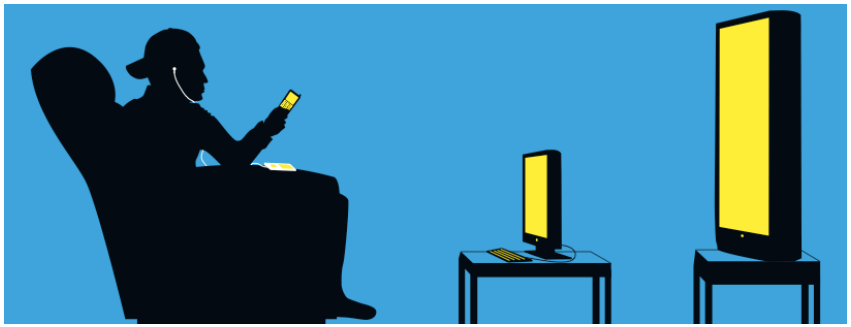
Roger Bohn
Jim Short
UC San Diego

http://hmi.ucsd.edu/howmuchinfo_research_report_consum.php

What is HMI?

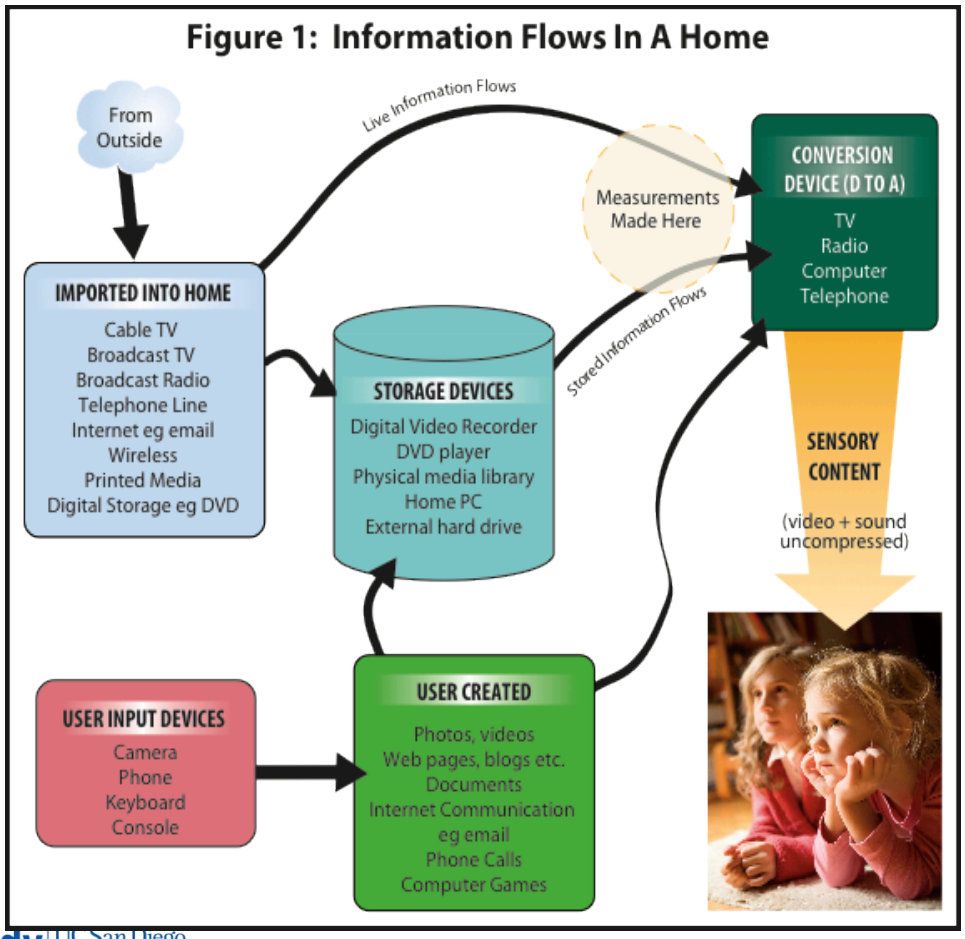
- A research program at UCSD and SDSC
- Our goal: create a census of the World's information
  - How much is there?
  - Of what types?
  - How is it created and where does it go?
- Measuring data and information
  - Inexact science
  - Assumptions and methods key

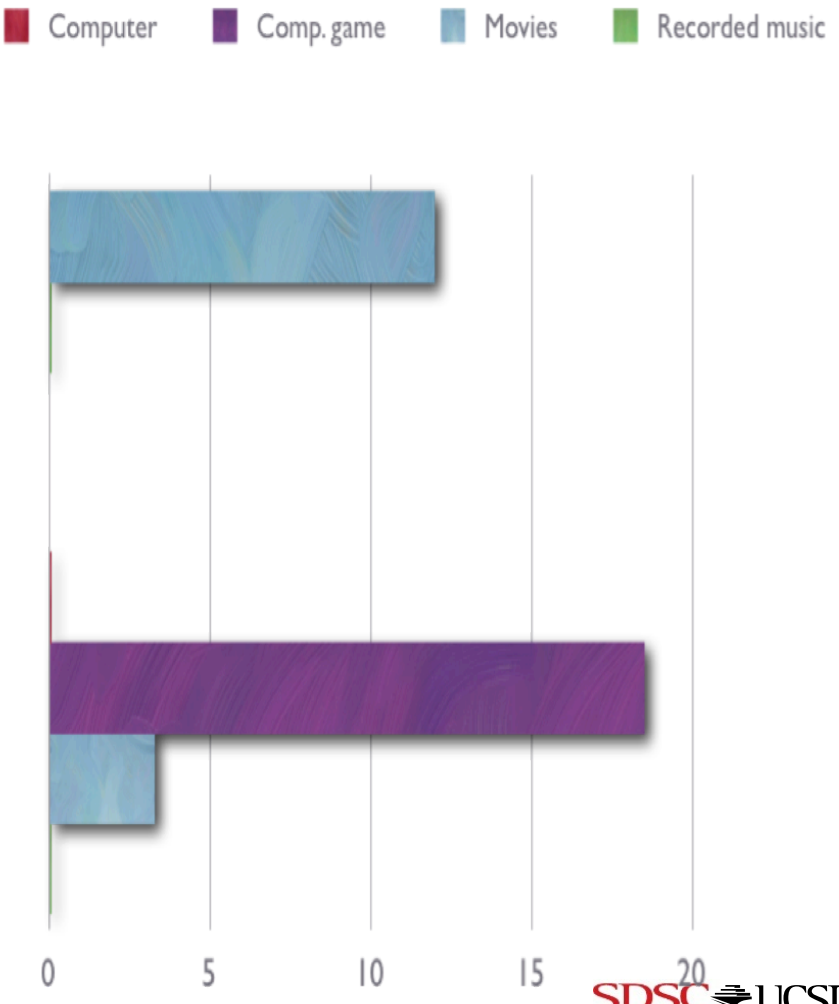
Consumers



23

Information to American consumers: 2008 average day





24

34 GB, 12 hours, 100,000 words

	Infoc in GB/day	Infow in words/ day	Hours per day
TV (incl. DVR, Internet, mobile)	12.0	44,342	4.9
Radio	0.1	8,315	2.2
Phone	0.01	5,269	0.7
Print	0.01	8,659	0.6
Computer	0.08	27,122	1.9
Computer games	18.5	2,459	0.9
Movies	3.3	198	0.03
Recorded music	0.08	1,112	0.45
Total	34.0	~100,000	~ 12


2K SPORTS
NBA 2K10
JORDAN
ORLANDO
2
24
14
30

26

Slow Growth: 5% CAGR

	1960 to 1980	1980 to 2008
$INFO_{Hours}$	3.9%	2.6%
$INFO_{Words}$	3.7%	3.0%
$INFO_{C}$ (bytes)	2.9%	**5.4%**
US Population	1.1%	1.0%
GDP $/capita	3.6%	2.9%

Source: HMI? 2009

Facebook, Twitter Stats

- Facebook
- Conventional Stats:

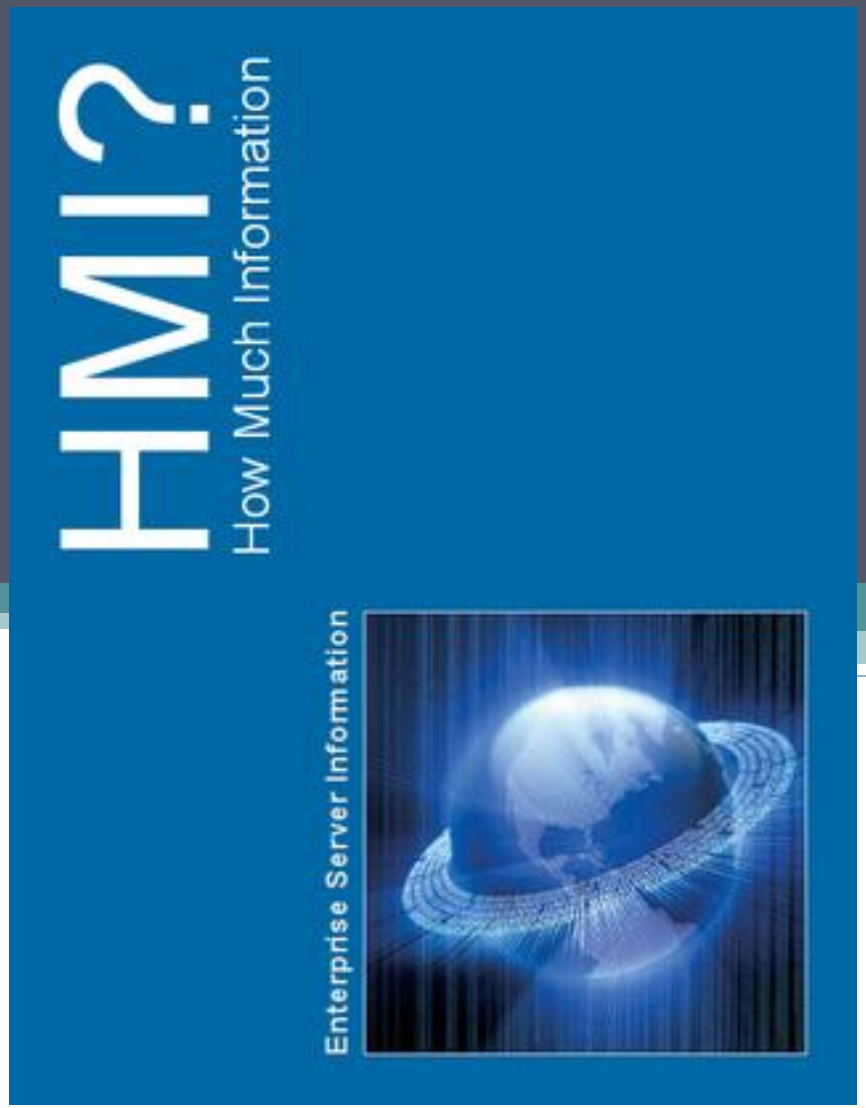
- 600,000,000+ users
- Half log in daily
- Average user: 130 friends
- Average session time: 23.3 mins per day
- Average user visits: 11.8 per month
- Core 18 to 24 year old segment growing 75% a year
- Over 250 million people access Facebook through their mobile phone

Twitter

Conventional Stats:

- 106,000,000 registered users
- 300,000 new sign up daily
- Average session time: 13.1 mins per day
- Average user visits: 10.2 per month
- 37% of registered users update their status through a mobile phone

How Much Enterprise Information Is There?



Findings from HMI? 2010

Jim Short
Roger Bohn
Chaitan Baru
UC San Diego and SDSC

Summary: How Many Bytes?

Table 1 World Server Information

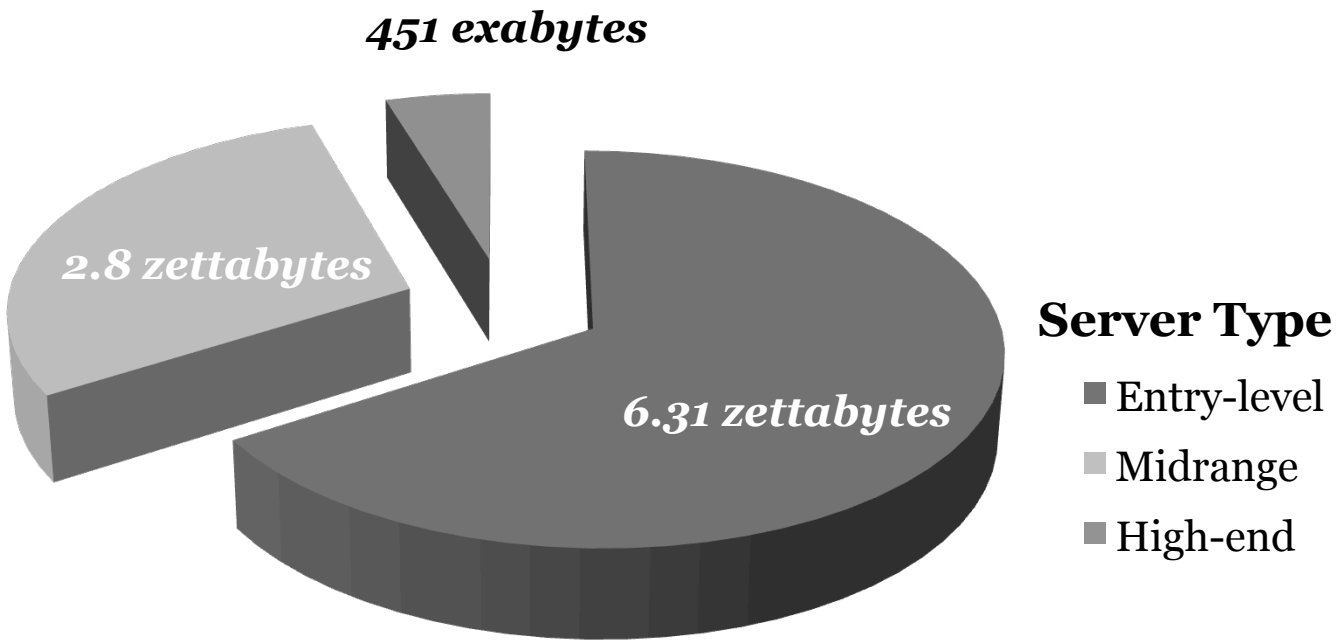
What is measured	World 2008 Total	Notes
Bytes processed plus delivered	9.57 zettabytes	
Bytes per worker per year per day	3.01 terabytes 12.0 gigabytes	3.18 billion workers in world labor force
Bytes per company per year	63.4 terabytes	151 million world businesses registered

`

Table 3: World Server Information by Server Class 2008

	Entry-level	Midrange	High-end	Total
Total Bytes by Server Class (in zettabytes)	6.31	2.80	.451	9.57

World Server Summary Information



IDC/EMC 2010 Total Digital Universe: 1.2 zettabytes in 2010
35 zettabytes by 2020

Lyman/Varian 2003: 18 exabytes of new information produced annually

Enterprise Information *Could Include*:

~~Data delivered to workers – screen-based~~

~~Data stored on storage media (what about redundancy?)~~

Data used (inputs & outputs - to inform, to process something, to take action)

~~Data in embedded processors in office & industrial machines~~

Our model:

"Information" = **Data** processed plus delivered for use

Measured as: **Bytes** processed plus delivered by servers

Servers = **World** installed base in 2008

Measuring "Work"

WORK = WORK Performed by Servers

We do not measure all workloads. But, our capacity measure accounts for all servers and we assume the workloads we do measure are representative of the whole

Measured By

Assumptions:

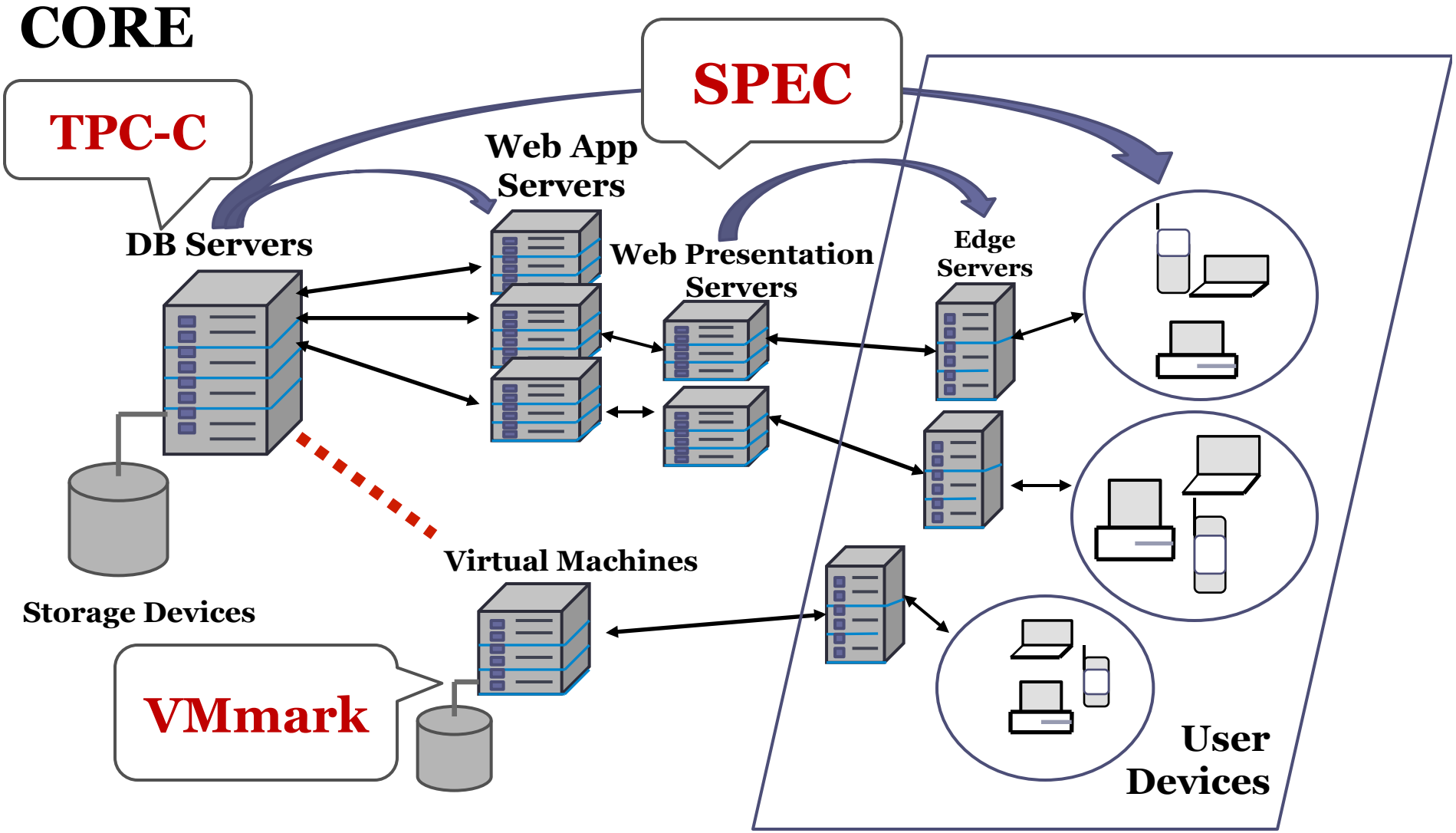
Simulated workloads represent actual workloads

Estimating max server processing capacity for multiple workloads yields meaningful upper limit

Industry Standard Server Benchmarks

Where Each Benchmark Simulates One or More Enterprise Workloads

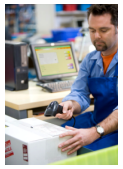
Measurement Points



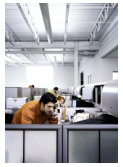
The TPC-C Simulated Workflow

"Workload" – online transaction processing; taking an order, entering it into a database application, interacting with the application to process and complete the transaction.

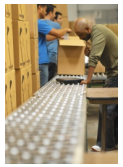
Transaction Mix


New Order (45%)


Payment (43%)


Order Status (5%)


Stock Level (5%)


Delivery (5%)

Client System

Mid-Tier Transaction Servers With Queuing Application

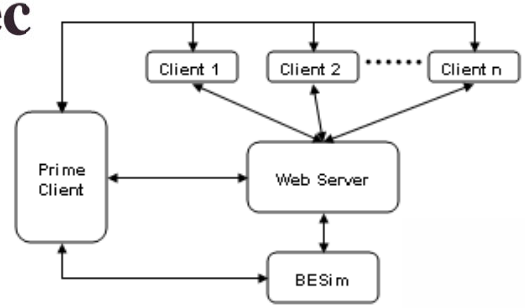
Database Server

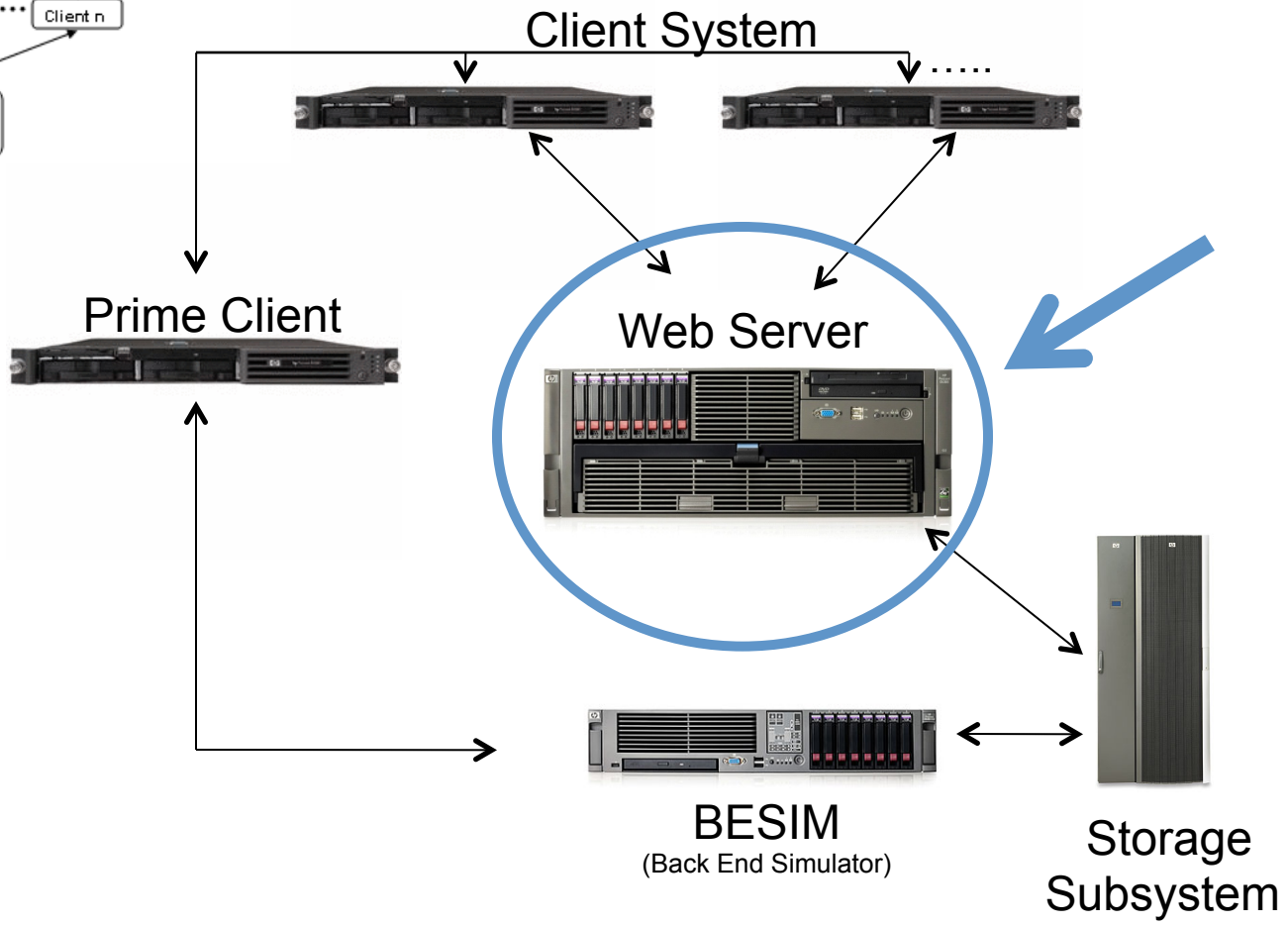
Storage Subsystem

Order of Magnitude: 95,000 tpmC "transactions per minute C" for a typical Volume server in 2008 (Dell 2900). In bytes, 15.4 billion bytes per minute

Source: Adapted from HP, "An overview of the TPC-C benchmark on HP ProLiant servers and server blades" August 2007

The SPECweb2005 Simulated Workflow



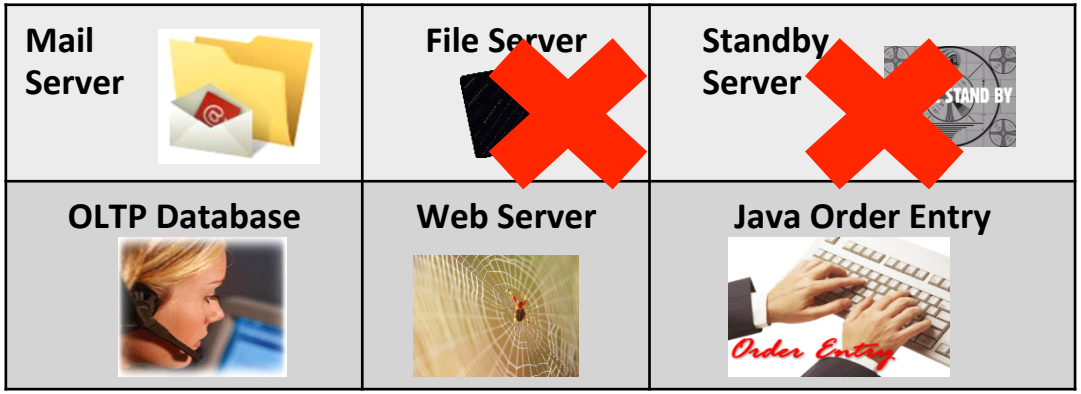


eCommerce

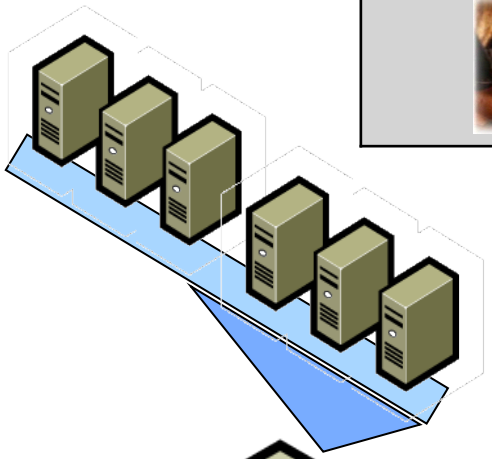
Request Type
index
search
browse
browse product line
product detail
customize1
customize2
customize3
cart
login
shipping
billing
confirm
Total

Order of Magnitude: 15,000 "SUS" (simultaneous user sessions) for a typical Volume server in 2008; 300,000 user requests per minute. In bytes, 9.6 billion bytes per minute.

VMmark Simulated Workloads



VMs



Test Server

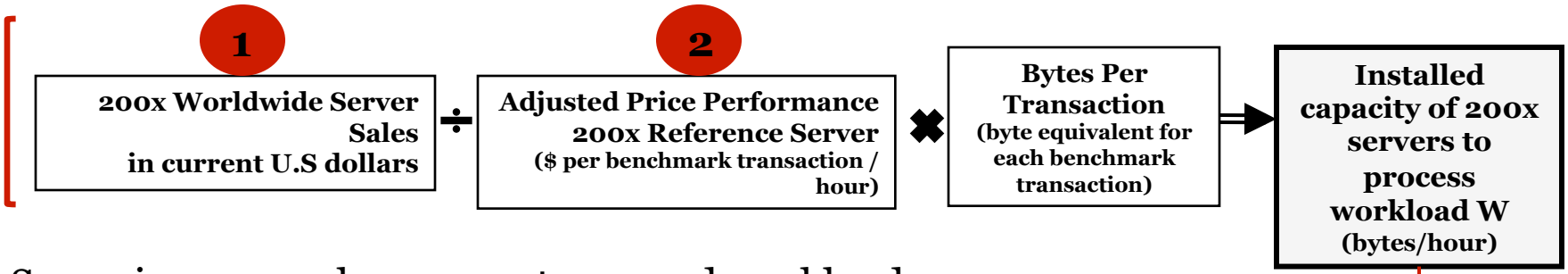
Six Virtual Machines = One VM Tile

Simulated Workloads:
Mail Server (Exchange mail server)
OLTP Database (MySQL simulation)
Web Server (SPECweb2005 design)
Java Order Entry (TPC-C like)

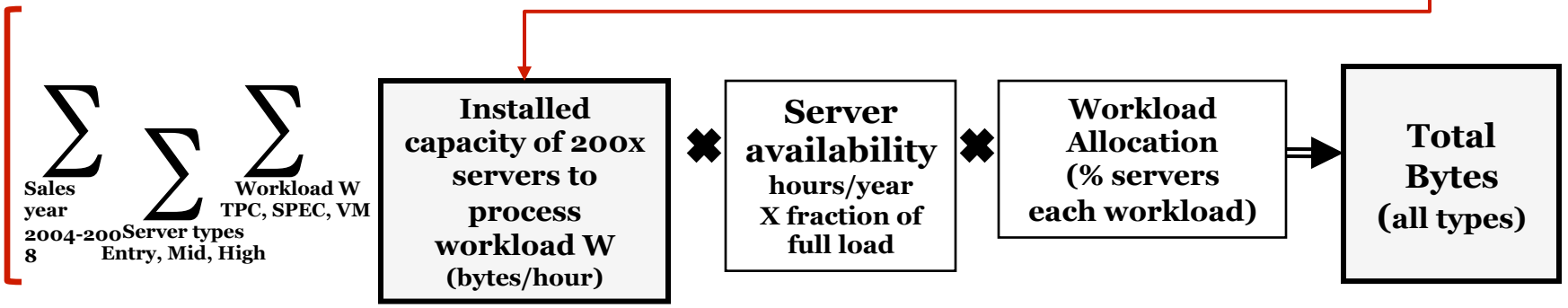
Order of Magnitude:
Mail Server – 1,500 actions per minute
Java Server – 17,500 new orders per minute
In bytes, a typical Volume server processes approximately 1.8 billion bytes per minute for the 4 workloads measured

Modeling Approach

1 By sales year 2004-2008
By server type (entry-level, midrange, high-end)
By workload **W** (TPC, SPEC, VMmark)



2 Summing over sales, server types and workloads



1

World Server Sales 2004-2008

Annual World Server Sales 2004-2008
Current U.S. Dollars (in billions)

Server Class	2004	2005	2006	2007	2008	Total
Entry-level	$24.4	$27.3	$28.5	$30.8	$29.3	$140.5
Midrange	$12.8	$12.8	$12.2	$12.6	$11.7	$62.3
High-end	$12.2	$11.6	$12.0	$11.6	$12.2	$59.9
Total	$49.5	$51.8	$52.8	$55.1	$53.3	$262.7

Source:
HMI? 2010. Data compiled from IDC Quarterly Server Tracking Reports, 2004-2009.

2

COST ANALYSIS

- We used the detailed TPC-C pricing sheets to break down the costs of the System Under Test (SUT) into the cost of the Server Hardware and other components
- To accord with our server sales numbers, we recalculated price performance using only the costs of the server hardware (current dollars)
- Further cost performance (storage, network) analyses are possible and planned

	PowerEdge 2900	TPC-C 5.9 TPC Pricing 1.2 Report Date June 16, 2008 Revision Date Availability Date June 16, 2008				
Description	**Part Number**	**Price Source**	**Unit Price**	**Qty**	**Extended Price**	**3 yr. Maint. Price**
Server Hardware						
PE2900,QC XEON X5440,2X6MB,2.83GZ	223-4506	1	1,253.00	1	$1,253.00	$578.00
& 2 Broadcom NICs						
32GB 667MHz(8x4GB),2R	311-6327	1	3,118.00	1	$3,118.00	
PERC6/i,Integrated	341-5699	1	$299.00	1	$299.00	
PERC6/E SAS RAID, 2X4 EXTERNAL	341-5842	1	$799.00	3	$2,397.00	
DELL E157FP,15 IN,15.0 VIS	320-5090	1	$189.00	1	$189.00	
				Subtotal	$7,256.00	$578.00
Server Storage						
PV MD1000,RACK,3U,15 BAY,LBZL	222-2299	1	2,480.00	6	$14,880.00	$9,888.00
SINGLE ENCL MGT MODULES, SAS/SATA	420-5927	1	$345.83	6	$2,074.98	
SAS Cable, 1M, MD1000	310-7082	1	$30.00	6	$180.00	
73GB,3GBPS,SAS,3.5IN,15K	341-3023	1	$299.00	98	$29,302.00	
42U Rack,CUST	340-4896	1	$239.99	1	$239.99	
				Subtotal	$46,676.97	$9,888.00
Server Software						
Oracle Database 11g Standard Edition One,Per Processor Unlimited Users, 3 years		2	$2,498.00	1	$2,498.00	
Windows Server 2003 Standard x64 Server	420-7118	1	$799.00	1	$799.00	
Microsoft Problem Resolution Services		3	$245.00	1		$245.00
Oracle Premium Support, 3 years		2	$1,099.00	3		$3,297.00
				Subtotal	$3,297.00	$3,542.00

1	$760.00	$472.00	
1	$599.00		
1	$448.00		
1	$0.00		
1	$59.00		
1	$99.00		
1	$149.00		
Subtotal	$2,114.00	$472.00	
1	$799.00		
1	$250.00		
Subtotal	$1,049.00	$0.00	
3	$3.93		
Subtotal	$3.93	$0.00	
16% discount	($8,967.52)		
Total	$51,429.38	$14,480.00	

*All hardware items from Dell(1) are discounted 16% based on total dollar volume of this configuration.

Price Source: 1=Dell, 2=Oracle,3=Microsoft,4=Kalron

Pricing may be verified by calling 1-800-BUY-DELL and referencing quote # 432979497 as a complex quote.

Audited by Lorna Livingtree, Performance Metrics Inc.

Three-Year Cost of Ownership:	**$65,910**	**USD**
TPC-C Throughput:	**97,083.53**	**tpmC**
Price/Performance:	**$0.68**	**tpmC/USD**

Prices used in TPC benchmarks reflect the actual prices a customer would pay for a one-time purchase of the stated components. Individually negotiated discounts are not permitted. Special prices based on assumptions about past or future purchases are not permitted. All discounts reflect standard pricing policies for the listed components. For complete details, see the pricing sections of the TPC benchmark specifications. If you find that the stated prices are not available according to these items, please inform the TPC at pricing@tpc.org.

Server Hardware: $7,256

Storage System: $46,677

SUT Test System: $65,910

World Server Capacities

ENTRY-LEVEL SERVERS	Reference Year	tpmC	Price/tpmC	Total System Cost	Server Hardware (SH) Cost	Adjusted SH_Price/tpmC	Core Capacity tpmC
Dell PowerEdge 2900	**2008**	**97,083**	**$0.68**	**$65,910**	**$7,256**	**$0.07**	**3.92E+11**
Dell PowerEdge 2900	2007	69,564	$0.91	$63,080	$11,658	$0.17	1.84E+11
Dell PowerEdge 2900	2006	65,833	$0.98	$64,512	$9,839	$0.15	1.91E+11
Dell PowerEdge 2800	2005	38,622	$0.99	$38,028	$10,771	$0.28	9.81E+10
Dell PowerEdge 2850	2004	26,410	$1.53	$40,170	$7,993	$0.30	8.08E+10
Estimated	2003					$0.33	7.34E+10

MIDRANGE SERVERS	Reference Year	tpmC	Price/tpmC	Total System Cost	Server Hardware (SH) Cost	Adjusted SH_Price/tpmC	Core Capacity tpmC
HP ProLiant DL585 G5	2008	402,234	$1.26	$502,836	$43,634	$0.11	1.08E+11
HP ProLiant DL580 G5	2007	407,079	$1.71	$694,335	$70,560	$0.17	7.33E+10
HP ProLiant ML370 G5	2006	240,737	$1.85	$443,443	$83,504	$0.35	3.53E+10
HP ProLiant DL585	2005	130,623	$2.80	$364,539	$87,372	$0.67	1.92E+10
HP ProLiant DL580 G2	2004	95,163	$2.93	$278,114	$72,151	$0.76	1.69E+10
Estimated	2003					$0.87	1.47E+10

HIGH-END SERVERS	Reference Year	tpmC	Price/tpmC	Total Price	Server Hardware (SH) Cost	Adjusted SH_Price/tpmC	Core Capacity tpmC
IBM System x3950 M2	2008	841,809	$3.46	$2,911,484	$1,020,576	$1.21	1.01E+10
IBM System p 570	2007	404,462	$3.50	$1,417,121	$625,499	$1.55	7.50E+09
IBM System p5 570	2006	203,440	$3.93	$799,990	$514,449	$2.53	4.76E+09
IBM eServer p5 570	2005	194,391	$5.62	$1,092,119	$527,839	$2.72	4.29E+09
IBM eServer p5 570	2004	371,044	$5.26	$1,951,215	$1,035,538	$2.79	4.39E+09
Estimated	2003					$3.35	3.66E+09

New Way To Measure Capacity

- Aggregate capacity over very different kinds of servers
- Assume companies spend dollars for server capacity efficiently
- Measure capital cost per unit of benchmarked performance

- Then we translate different benchmarks into a common unit: bytes

Business Considerations

Cloud Computing In the Era of Big Data

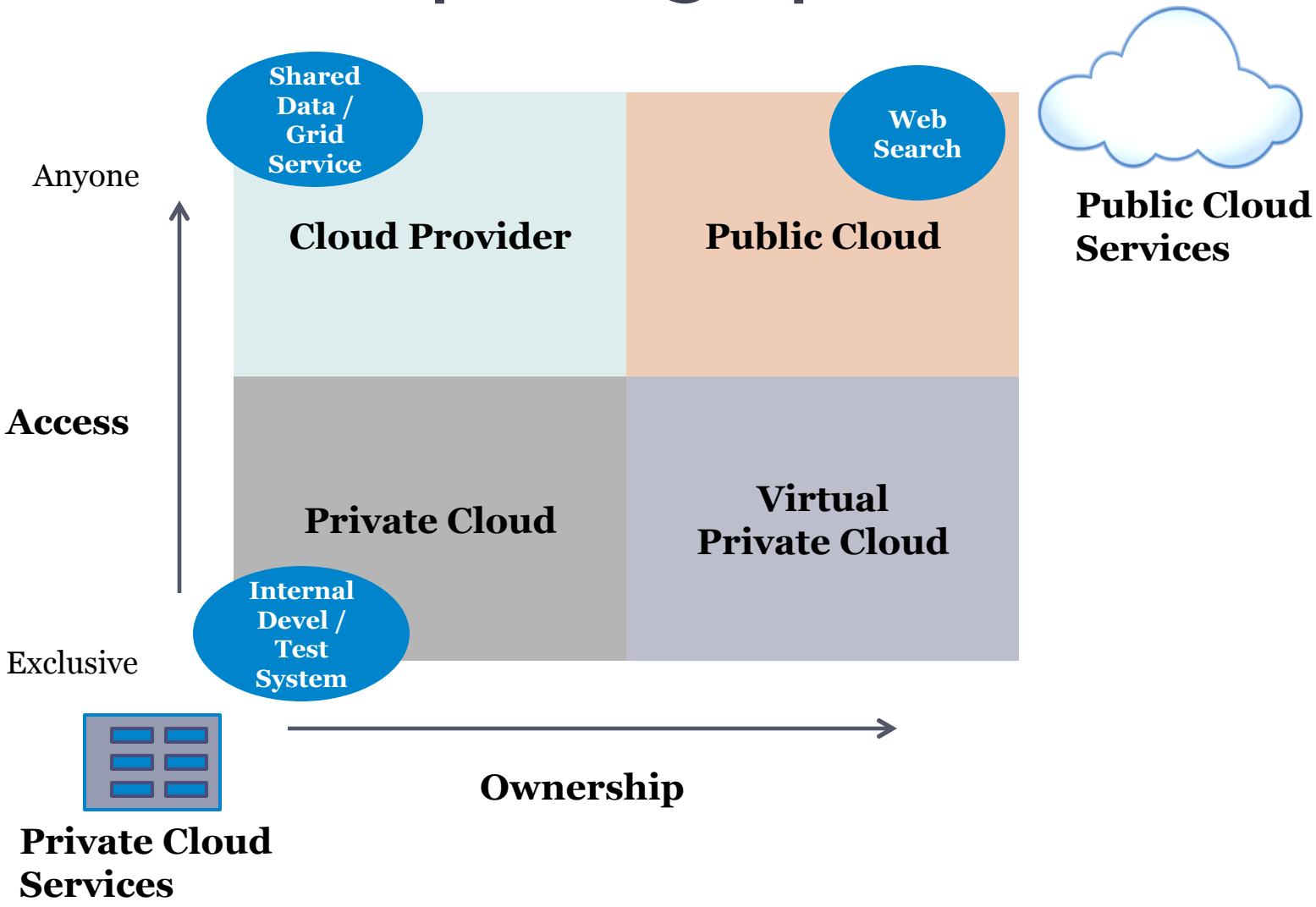
Cloud Potential Benefits

- C-level execs want to "press a button" to get more IT resources to meet dynamic needs
- CAPEX to variable OPEX
- Usage-based billing
- Launch speed and application agility
- Ease of maintenance and resource flexibility (free up IT staff to focus on business drivers)
- Energy efficiency

Cloud Concerns

- Suitability for needs – defining appropriate workloads and applications
- Security and privacy (legal stays with owner)
- Availability, latency, reliability, performance, lack of reporting
- Data ownership and control (visibility, access)
- Interoperability (no standards)
- Vendor management (contracting, charge-back and sharing)

Cloud Computing Spectrum



What Are People Doing Now?

- Prototyping / Proof of Concept
  - Setting up private clouds to test cloud performance
    - Web content and applications (web analytics)
    - Email and web collaboration (for large user bases)
    - Software as a Service (SAAS) – peripheral processing, while keeping core functions (ERP) in the datacenter
    - Test and development – rapid provisioning, unpredictable workloads
    - HPC – economics of scale for very large data queries and massively parallel workloads

Cloud Computing Cost Models

ebizQ From http://www.ebizq.net/blogs/enterprise by Dion Hinchcliffe

Cloud Top 10

- #1 Ownership
  - Who in your enterprise "owns" cloud? Who is responsible for it? Is this agreed?
- #2 Cost
  - Who pays for cloud? (innovation is not cheap) Will users pay? Are you already using chargeback?
- # 3 Regulatory / Legal
  - Is corporate legal involved and do they "get it"? You have responsibility for your data, not your services provider. Does legal have what they need to be effective partners in contracting and managing cloud service providers?

Cloud Top 10

- #4 Leveraging Cloud
  - Is your company aggressive in rethinking its IT architecture with cloud? Will the potential benefits of be realized through effective implementation of visible cloud strategies?
- #5 Shared Infrastructure
  - Underlying network and storage resources need to be shared.
- #6 Scalable
  - An effective cloud solution has to be scalable.
- # 7 Multi-Tenancy
  - Strong multi-tenancy involves extensive use of VLANs to isolate network traffic between different zones in the cloud.

Cloud Top 10

- #8 Chargeback
  - IT organizations must be able to create effective and accurate chargeback capabilities.
- #9 Define, Measure, Track Performance & Costs
  - Since there are no standard tracking mechanisms used across cloud services, IT administrators will need to build or find ones appropriate to their circumstances
- #10 Implement Bottom Up, Manage Top Down
  - Cloud involves a major change in how business users access and use IT resources.